Université Lille Nord de France
Pôle de Recherche et d'Enseignement Supérieur

Ecole doctorale régionale Sciences Pour l'Ingénieur Lille Nord-de-France - 072

Titre : Bornes sur le risque en apprentissage des réseaux de neurones

Directeur de thèse : Nicolas Wicker
E-mail : nicolas.wicker@univ-lille.fr

Co-directeur de thèse : Philippe Heinrich
E-mail : philippe.heinrich@univ-lille.fr

Laboratoire : Paul Painlevé

Équipe : Probabilités et statistiques

Descriptif :
Le but de la thèse est d'affiner les bornes connues sur les mesures de capacité (complexité de Rademacher, nombres de recouvrement) dans le cadre des réseaux de neurones. Ce qui motive un tel sujet est que l'on cherche à majorer, avec une grande probabilité, un risque théorique R(f) sur une fonction f parcourant une classe de fonctions F à partir de son pendant observé $\hat{R}(f)$ sur un m-échantillon d'apprentissage $S = \{x_1,...,x_m\}$ et d'un terme $R_m(F)$ mesurant la capacité de la classe F à capturer le modèle adéquat. Cette mesure de capacité n'est pas estimée de façon satisfaisante dans le cadre des réseaux de neurones.

Dans un cadre plus général [Mohri et al., 2012], l'inégalité typique indique qu'avec probabilité supérieure à 1 − δ, on a pour toute fonction f de F,

$$R(f) \leqslant \hat{R}(f) + R_m(f) + \sqrt{\frac{\log(1/\delta)}{2m}}$$

où le terme $R_m(F)$ est ici la *complexité de Rademacher* de la classe de fonction F :

$$R_m(F) = \mathbf{E}_{S \sim D_m}\left(\hat{R}_S(F)\right).$$

Cette complexité est l'espérance sur le m-échantillon S selon une loi donnée D_m, d'une complexité de Rademacher $\hat{R}_S(F)$ que l'on peut calculer sur S :

$$\hat{R}_S(F) = \mathbf{E}_{\sigma \in \{-1,+1\}^m}\left(\sup_{f \in F}\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i f(x_i)\right]\right)$$

L'estimation de $\hat{R}_S(F)$ existante dans la littérature [Anthony, 2005, Neyshabur et al., 2015, Bartlett et al., 2017] n'est pas vraiment satisfaisante en ce qui concerne les réseaux de neurones. L'objectif est donc ici de revisiter les résultats connus et d'en tirer des raffinements. Une première piste pourrait partir de la constatation que lorsqu'on décompose la complexité de Rademacher observée suivant la dernière couche cachée, à savoir

$$\hat{R}_S(F)=\mathbf{E}_{\sigma \in \{-1,+1\}^m}\left(\sup_{f_j\in F_j,\alpha_j}\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i \sum_{j=1}^{k} \alpha_j f(x_j)\right]\right)$$

dans la plupart des preuves qui nous sont connues, la dépendance entre les classes de fonctions $F_1,\ldots,F_k$ représentant les réseaux de neurones obtenues en tronquant à la couche précédente n'est pas prise en compte. Si elle l'était, les bornes seraient certainement plus fines. Une approche par les nombres de recouvrement faisant appel aux techniques de chaînage de Dudley (ou plus généralement Talagrand) est également envisagée.

Références :

1. M. Anthony. *Connections between neural networks and boolean functions,* 2005.
2. P.L. Bartlett, D.J. Foster, and M. Telgarsky. *Spectrally-normalised margin bounds for neural networks*. In 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017.
3. M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. MIT Press, 2012.
4. B. Neyshabur, R. Tomioka, and N. Srebro. *Norm-based capacity control in neural networks*. In JMLR : Workshop and conference proceedings, volume 40, pages 1–26, 2015.
5. M. Talagrand. *Upper and Lower Bounds for Stochastic Process*, Springer, 2014.